

PAPER

How to write a systematic review of reasons

Daniel Strech,¹ Neema Sofaer²

► Additional materials are published online only. To view these files please visit the journal online (<http://jme.bmj.com/content/38/2.toc>).

¹Assistant professor, Hannover Medical School, CELLS - Centre for Ethics and Law in the Life Sciences, Institute of History, Ethics and Philosophy, Hannover, Germany

²Wellcome Trust Research Fellow, Centre of Medical Law and Ethics, School of Law, King's College London, London, UK

Correspondence to

Professor Daniel Strech, Hannover Medical School, CELLS - Centre for Ethics and Law in the Life Sciences, Institute of History, Ethics and Philosophy, Carl Neuberg Strasse 1, 30625 Hannover, Germany; strech.daniel@mh-hannover.de

Received 8 July 2011

Revised 24 August 2011

Accepted 1 September 2011

Published Online First

11 November 2011

ABSTRACT

Systematic reviews, which were developed to improve policy-making and clinical decision-making, answer an empirical question based on a minimally biased appraisal of all the relevant empirical studies. A model is presented here for writing systematic reviews of argument-based literature: literature that uses arguments to address conceptual questions, such as whether abortion is morally permissible or whether research participants should be legally entitled to compensation for sustaining research-related injury. Such reviews aim to improve ethically relevant decisions in healthcare, research or policy. They are better tools than informal reviews or samples of literature with respect to the identification of the reasons relevant to a conceptual question, and they enable the setting of agendas for conceptual and empirical research necessary for sound policy-making. This model comprises prescriptions for writing the systematic review's review question and eligibility criteria, the identification of the relevant literature, the type of data to extract on reasons and publications, and the derivation and presentation of results. This paper explains how to adapt the model to the review question, literature reviewed and intended readers, who may be decision-makers or academics. Obstacles to the model's application are described and addressed, and limitations of the model are identified.

Systematic reviews traditionally answer an empirical question based on an unbiased assessment of all the empirical studies that address it. Such reviews emerged in the 1970s in social science and were developed to a high level of sophistication in medicine and epidemiology. The literature that addresses questions in these fields is large and of varying quality; some is difficult to retrieve. Policy-makers and professionals in healthcare and research may lack the time or skills to collect, appraise and synthesise all the relevant literature. Systematic reviews undertake this substantial task and answer the question in a form accessible to decision-makers.¹

The process of a systematic review comprises four steps (box 1). The PRISMA statement gives standards for executing these.² Some of the standards ensure that the process is transparent, enabling readers to assess its adequacy and to reproduce it. The point of the process's systematic nature is to collect all the relevant literature and to minimise bias in characterising it.

The genre was subsequently transferred to qualitative research and the overlapping and burgeoning field of empirical bioethics, which uses empirical (frequently qualitative) studies to answer empirical questions relevant to bioethics.^{3–7} The intent of all these applications has been to leverage the existing literature to improve decision-making.

Some have recently advocated applying the genre to argument-based literature in clinical and research ethics, and in bioethics generally, again to improve decision-making, and there have been two such applications.^{8–9} Argument-based literature uses arguments to address conceptual questions, such as whether abortion is ever morally permissible or whether research participants should be legally entitled to compensation for sustaining research-related injury. We agree with McCullough *et al*⁹ that clinicians could benefit from systematic reviews of clinical ethics literature. However, as we argue at length elsewhere, there is a need for a much more sweeping adaptation of the systematic review technique, and engagement with the many technical and conceptual issues, for such reviews to accomplish their goals in clinical and policy decision-making.¹⁰

In more detail: with respect to the first step of writing a systematic review, McCullough *et al*⁹ propose that a systematic review of clinical ethics literature should address an ethical question. Their review of a seven-article literature addresses the following question:

'In patients with mental disorders (schizophrenia, dementia), is use of concealed medications in food or drink, rather than prescribing medications in the usual way or forcibly administering them, ethically justifiable?'

It has the same form as the review question of a traditional systematic review in medicine or epidemiology (the so-called PICO scheme): it mentions a population, intervention, comparison and outcome.¹¹ The only change that occurs in the transfer from the medical to the bioethics literature is in the outcome: from a physical outcome, such as increased mortality, to an ethical outcome, (here) ethical justifiability. To answer this question, in step 3 (see box 1) McCullough *et al*⁹ extract, from each publication included in the review, the publication's all-things-considered conclusion and a single numerical score that reflects 'the adequacy of the ethical analysis and argument' (p. 67) from which this conclusion was drawn. Regarding step 4, they consider that the answer to the review question is the answer most commonly given by the included publications, when greater weight is given to answers based on higher-scoring reasoning.⁹ Importantly, they do not propose in step 3 the systematic extraction and synthesis of information on the reasons given when discussing the ethical question and how they were used. We call their outline model for writing systematic reviews a systematic review of (quality-weighted) conclusions.



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://jme.bmj.com/site/about/unlocked.xhtml>

Box 1 Four steps for writing a systematic review

1. Formulate the review question and eligibility criteria.
2. Identify all of the literature that meets the eligibility criteria.
3. Extract and synthesise data.
4. Derive and present results: the answer to the review question.

As we argue elsewhere, we must reject McCullough *et al*'s measure of 'the adequacy of the ethical analysis and argument' (p. 67).¹⁰ Furthermore, while it might be possible to replace it with a suitable measure, a systematic review that answers an ethical question may mislead decision-makers when the literature reviewed is incomplete or inadequate. In these cases, the literature's answer to the review question places no burden of proof on those who disagree. Of course, when an empirical literature is inadequate, its answer will also be potentially misleading and uninteresting; the correct safeguard in both cases (inadequate empirical and reason-based literatures) is for the review to conclude only that further research is needed to answer the question. To date, however, the assessment of the quality of reasons and of argument-based literature is much less standardised than, for example, the assessment of the quality of clinical trials and the literature that reports their results. Bioethicists as well as clinical and policy decision-makers are less likely, we surmise, to understand the significance of limitations in reasoning than in study design. A McCullough model systematic review, insofar as it is a systematic review of (quality-weighted) conclusions, also has normative problems: it may mislead when there are mutually incompatible, but maximally informed and individually reasonable, answers to the ethical question, or when different weightings of the reasons (as may be appropriate in dissimilar contexts) support different answers.

Our alternative model for writing systematic reviews of argument-based literature proposes that the review question should be not an ethical question but the factual question of which reasons have been given when discussing the ethical question and how they have been used. Our pilot systematic review addressed the question:

'Which reasons have been given for the views that former participants in a drug trial should, or need not, be ensured post-trial access (PTA) to the trial drug?'⁸

We call such systematic reviews of argument-based literature systematic reviews of reasons. Our first such review identified and presented the reasons given in all their variants, and their alleged implications, and whether authors accepted or rejected the reasons.

Such detailed information on reasons is crucial for decision-makers and philosophers. Both need to identify all the strong (and thus relevant) reasons and their implications for the relevant decision or ethical question. A review of reasons cannot guarantee to accomplish this for them: the reviewed literature may omit relevant reasons or be wrong about which reasons are relevant. However, such a review reduces the risk of neglecting relevant reasons, or interpretations thereof, or their possible implications. A systematic review of reasons is likely to reveal a greater range of such information than the informal reviews of reasons that are usual in bioethics and philosophy, which sample literature using unsystematic, undocumented search methods to the unspecified point at which it seems to the author (often the only author) that no relevant new reasons emerge. The difference is likely to be marked when a literature is large, fragmented across disciplines

and literary genres, and indexed in databases inadequately and inconsistently, as bioethics literatures often are.⁸

Furthermore, systematic reviews of reasons also help to improve argument-based bioethics by identifying gaps such as reasons that have been presented only inadequately, or factual claims that need testing. Our systematic review showed differences between publications on the cost, legality and logistics of ensuring PTA, and suggested that many factual claims were not evidence based.⁸ So, we surmise, reviews of reasons suggest areas for further empirical and philosophical research to social scientists, economists, lawyers and philosophers, which would improve the information base for decision-making. Again, the relevance of the systematic nature of the review is that a greater variety of reasons is likely to be identified. However, the review itself neither involves nor replaces the critical analysis and weighting of reasons.

While we argue elsewhere in more detail why and when bioethics need such systematic reviews of reasons,¹⁰ the literature still lacks a comprehensive explanation and justification of the different steps of a systematic review of reasons. Here we present our model for writing systematic reviews of reasons, which we have structured according to the four steps in box 1, but differs from models for writing systematic reviews in epidemiology or social science literature.^{1 7 12} While we illustrate it using our first systematic review of reasons,⁸ it applies to all argument-based literature. The appendix (available online only) explains how we developed the model, both to justify its appropriateness to our particular systematic review and to explain how to adapt the model to new review questions or literatures.

MODEL FOR WRITING SYSTEMATIC REVIEWS OF REASONS

Formulate the review question and eligibility criteria

A tentative general form of review question is:

'Which reasons have been given for the views that action or policy X is, or is not, permissible (alternatively: required/forbidden)?'

As mentioned above, our pilot systematic review addressed the question:

'Which reasons have been given for the views that former participants in a drug trial should, or need not, be ensured post-trial access (PTA) to the trial drug?'⁸

It may be necessary to specify whether the requirement is moral or legal. Publications arguing that X is not required may hold that X is permissible or forbidden; as we note later, the analysis should be sufficiently sensitive to distinguish between these positions.

The eligibility criteria should identify all and only publications that include the reasons mentioned by the review question. For example, our eligibility criteria were:

- '... a publication, e.g. article, [should be included] if and only if:
1. It included a reason why PTA should or need not be provided;
 2. The PTA was for former participants in a drug trial;
 3. The PTA was to a drug tested in the trial; and
 4. The publication was a peer-reviewed, published academic article or book; national-level report or working paper; or PhD thesis.'

It will sometimes be necessary, as in our case, to explain how to interpret the criteria, or justify why they were chosen. Criteria for including or excluding publications based on language or ranges of publication dates will need to be explicitly stated and justified.

Identify all of the literature that meets the eligibility criteria

Databases and search techniques should be selected with the aim of retrieving all available literature meeting the eligibility criteria.

We used databases in science/medicine (Medline, LocatorPlus), law (Westlaw International) and ethics (ETHXWeb, JSTOR, Euroethics, Endebit) and thesis databases (Ethos-Beta Electronic Theses Online Service, WorldCat Dissertations). We recommend that, as in our review's case, the choice of databases should be guided by experts such as reference librarians.

Database-specific search strings will be needed. For each database, the controlled vocabulary (eg, Medline's MeSH terms) should first be examined to determine whether it contains the review question's keywords, for example, (in our case) PTA. If not, the use of non-controlled vocabulary will be necessary: one should identify some publications that meet the eligibility criteria; then, one should identify the database-specific controlled vocabulary used to index these publications, and keywords relevant to the review question, for example, "post-trial follow-up", and classify the resulting terms by content. Often a mixture of controlled and non-controlled vocabulary can help to adjust the sensitivity and specificity of search strings.^{8 13} One should next join terms in the same content class by 'OR' and all (alternatively, some) of the resulting strings by 'AND' to form database-specific search strings. (Table 1, gives key strings that we used to search Medline; see also reference⁹). One should also search as systematically as possible for relevant reports and books, hand-search their tables of contents and indexes, and examine the in-text references, footnotes and bibliographies of qualifying publications for further publications that possibly meet the eligibility criteria.^{7 14} The contents of electronic books can already be searched systematically and quickly by the search functions of PDF software. Further research is needed to construct systematic and efficient searches for books and within print-only books. Technologies such as the optical character recognition used, for example, by Google Books (<http://books.google.com>) could provide further opportunities to search content within books systematically; the way in which an electronic book has been prepared greatly affects its searchability.

When time constraints limit the inclusiveness of the search, authors should acknowledge this and explain why the search is nonetheless valid. Reference management software can be used to record searches and to quantify the overlap between searches and between databases.

A crucial step is the process used to determine which of the publications initially retrieved (in our case, 2039 publications) meet the eligibility criteria (in our case, 75 publications).⁸ Review authors should be aware that not all the literature that presents reasons relevant for the review question is presented as ethical literature.

Reviewers should work independently through the list of retrieved publications to exclude those that seem irrelevant based on their title, abstract and controlled vocabulary. Then,

discrepancies between the two (or more) resulting lists (depending on the number of reviewers) should be jointly resolved to create a single list of publications (ours had 146 publications) that both authors consider possibly meet the eligibility criteria. Each author should next read the full text of every listed publication. A publication should be included if, and only if, both authors agree that it meets the eligibility criteria. If they cannot reach agreement, an independent person should act as tie-breaker to enable the review process to continue; however, it is important to document, for example in an appendix, the grounds for the disagreement.

The written systematic review should include a flow chart that describes the selection of the publications included in the review. Together with a verbal description of the search strategy, this enables readers to reproduce the search and assess the likelihood that the review included all the qualifying publications. When the search was complex, inclusion of a list of databases searched with the database-specific search strings helps make the search reproducible. For examples of each, see reference.⁸

There is more to a systematic review of reasons than a search that seeks to be comprehensive, as the following sections explain.

Extract and synthesise data

We distinguish here between a reason mention (or mention), a reason expressed by a specific passage, from a reason type, a type of reason which may have different mentions in different publications.

To achieve a more comprehensive and less biased overview of reasons than an informal review, it is important to extract data on each mention and on the publication itself. The extraction and synthesis of reason types goes beyond the simple copying of text passages given in the original literature, involving several more interpretive tasks. First, a text passage needs to be identified as one that addresses a reason. Second, types of reason need to be generated based on these text passages: narrow types and broad types, which include the narrow ones. Different methods have been developed in qualitative research on how to develop broad and narrow types: thematic analysis, meta-ethnography, content analysis.^{3 15 16} These methods differ in how much weight they give to descriptive or interpretive tasks in the analysis and comparison of text passages.^{3 15} With respect to the extraction and synthesis of reasons, all qualitative research aims to compare text passages that mention reasons across papers and to match reason mentions from one paper with reason mentions from another, ensuring that a reason type captures similar reason mentions from different papers. Qualitative research also involves developing a hierarchy of narrow and broad codes.

Table 2, which should be read next, summarises the data we recommend should be extracted. Subsequent sections identify and address additional obstacles to, or limitations of, the extraction.

Table 1 Key Medline search strings used in⁸

Focus	Search string
Ethics	((("Ethics"[Mesh] OR "Human Rights"[Mesh] OR "ethics"[Subheading])) AND (((((((((((("Health Services Accessibility"[Mesh]) OR ("Continuity of Patient Care"[Mesh])) OR (("Drugs, Investigational/supply and distribution"[Mesh])) OR ((post-trial provision))) OR ((post-trial obligations))) OR ((post-trial access))) OR ((post-trial benefits))) OR ((post-trial responsibility))) OR ((follow-up))) AND (((("Clinical Trials as Topic"[Mesh]) OR ("Biomedical research"[Mesh])) OR ("Human experimentation"[Mesh])))))))
Developing countries	((((((("Vulnerable Populations"[MeSH Terms]) OR ("Developing Countries"[Mesh])) AND (((("Patient Advocacy"[MeSH Terms]) OR ("Ethics"[Mesh] OR "Human Rights"[Mesh] OR "ethics"[Subheading]))) AND (((((((("clinical trials as topic"[Mesh]) OR ("Human Experimentation"[Mesh])) OR ("Biomedical Research"[MeSH Terms])) OR ("Drugs, Investigational/supply and distribution"[Mesh])) OR ("Research/organization and administration"[Mesh])))))))

While some of the parentheses in the search strings are logically redundant, the search can be repeated in Medline only by using the strings as written here.

Table 2 Data to extract from publications included in the systematic review

Type of data	How to code the data	Comments
Data to extract on each reason mention		
Content	Assign to each reason mention: <ul style="list-style-type: none"> ▶ A broad reason type, for example, avoid exploitation ▶ A narrow reason type, for example, avoid exploiting research participants or avoid exploiting the host country. 	To minimise bias, it is best to assign reason types to a passage based only on the words in the passage, because different reviewers can agree on which words are used. An exception is when publications use the same word, for example "reciprocity" to express different reasons, such as reciprocity and distributive justice. See the text below this table (Content). One should think about the review's purpose when deciding how broad to make reason types, whether types should be mutually exclusive, and whether mentions should be assigned more than one narrow (or broad) reason type. For example, if one purpose is to prompt examination of the relations between reasons that are considered distinct in the literature, different mentions should be assigned to different types whenever the suspicion arises that the types may differ. Whenever possible, reason types should be given short, self-explanatory names in order to keep lists of reasons accessible to decision-makers.
Alleged implications	Assign to each reason mention a number code depending on whether it is, for example, <ol style="list-style-type: none"> 1. allegedly for ensuring PTA to the trial drug, or 2. allegedly against ensuring this, or 3. the mention is claimed to have implications for ensuring PTA, but the implications are unspecified or unclear. 	Reasons against ensuring PTA, for example, may include reasons why PTA need not, or should not be provided. Most reviews will need to extract data that reflects such a distinction. We did this by extracting from each reason mention the conclusion drawn from the reason mention, as distinct from the all-things-considered conclusion drawn by the publication (see below).
Optional: person(s) expressing attitude to reason mention	Code each reason mention depending on whether the person expressing attitude to reason is <ol style="list-style-type: none"> 1. The author, or 2. Another party whose view the author reports If 2, consider recording the other party's name if given, and any relevant reference.	Extracting mentions to which other parties express an attitude allows reasons to be collected that are not directly published in the literature reviewed. Recording the other party's name and any relevant reference enables identification of the publications that are more frequently discussed, endorsed or rejected.
Attitude taken	Assign a number code to the attitude taken to the reason mention according to whether <ol style="list-style-type: none"> 1. the reason is always endorsed 2. the reason is sometimes endorsed 3. it is unclear whether the reason is endorsed or rejected 4. the reason is rejected. Example: assign 2 to a reason mention coded as reciprocity when the author asserts that reciprocity supports giving PTA in some contexts but in other contexts fails to apply.	Where the author's attitude to a reason changes in the course of a publication, context and order should be used to decide which attitude is 'authoritative'. A limitation is that rejected reasons receive the same attitude code, irrespective of whether they are considered irrelevant or based on incorrect factual or moral claims. The review's purpose should be considered when deciding how to balance removal of these limitations against keeping data accessible.
Any conclusion drawn from the reason mention	For each reason mention, note the conclusion drawn from it (as distinct from the all-things considered conclusion that the publication draws based on all the reasons it considers). Consider developing codes for conclusions that capture parts of the conclusion. Our review used codes to capture: <ul style="list-style-type: none"> ▶ The PTA-related action, for example funding PTA, making a pre-trial plan to ensure PTA ▶ Whether the action is said to be permissible, forbidden or required ▶ The agent held responsible for the PTA-related action (eg, researchers). 	It is particularly important to extract the conclusion drawn from the reason mention when mentions of the same reason differ as to the conclusion drawn. In the literature we reviewed, some conclusions drawn from reasons concerned whether there are moral obligations to ensure PTA, others whether it should be legally required. Reviewers will need to decide how narrow conclusion codes should be. There is a trade-off between ensuring that aggregate statistics are accessible, which requires broader codes, and meaningful. On the one hand, the latter tends to favour narrow codes but, on the other hand, if they become too narrow then the aggregate statistics lose their meaning.
Optional: reference	For each reason, consider noting whether an appropriate reference is given where one is necessary. Criteria will be needed. We considered references necessary except when reasons occur in the titles or abstract, or were clearly original when published. Code on the basis of whether the reason falls under the scope of an appropriate reference, rather than on whether there is a reference in the quotation extracted.	While collection of data on references is optional, it is particularly useful when writing a review that will enable academics to understand the state of a literature, as absence of appropriate references suggests (as in the case of our review) that some proponents of a view are ignorant of relevant publications. Be sensitive to different referencing conventions, and exercise caution when assessing the originality of reasons. Include appropriate disclosures in the limitations section.
Data to extract on the publication		
Any conclusion drawn by publication from all the reasons considered	Same as for conclusion drawn from the reason, except that the conclusion extracted is the all-things-considered conclusion that the publication draws on the basis of all the reasons it considers, as distinct from an interim conclusion that it draws from just one of these reasons.	Same as the comments on extracting conclusion drawn from the reason mention (see above).
Publication type	Collect various data on the publications included in the systematic review. For example: <ul style="list-style-type: none"> ▶ Broad (or narrow) type of publication for example, article, monograph (an example of a narrow type of publication is a policy review) ▶ Any restrictions on scope (eg, to HIV research) ▶ Reference Consider developing a measure for, and collecting data on, authors' conflict of interest	When possible, give eligibility criteria for types. In our first systematic review of reasons we also collected data on: <ul style="list-style-type: none"> ▶ Whether PTA to trial drug is publication's key topic (Y/N) ▶ Whether the publication's PTA content exclusively concerns research conducted by resource-rich country sponsors in resource-poor or middle-income countries (Y/N)

PTA, post-trial access.

Content

Assigning reason types on the basis of concepts

Systematic reviews of reasons aim to minimise bias in identifying and subsequently presenting reasons. However, systematic reviews are not free of bias. The risk of bias within systematic reviews of reasons increases when reviewers need to assign reason types to a reason mention based partly on the concept the mention expresses. Furthermore, whether types are assigned based on words or concepts, reviewers may disagree on how to divide a broad type of reason into narrow types. For example, we found competing ways of dividing the concept of reciprocity. One way was in terms of who needs to reciprocate and to whom (eg, reciprocity from society to research participants). The other, overlapping way was in terms of the benefits participants provide that might be thought to give rise to the reciprocal obligation (eg, reciprocity in return for assuming the risk of participation in research). Both minor codes would apply to a passage claiming that society should ensure PTA to participants in return for participants' assumption of the risks of research. In both the assignment of broad and narrow types, even if reviewers agree how to assign types, a different team might assign differently.

We recommend that:

1. There should be at least two reviewers. They should assign reason types to mentions independently, and identify, discuss and resolve discrepancies. When disagreement persists, the underlying reasoning should be stated and an independent person should be asked to break the tie.
2. The analysis should not imply greater precision than exists in the literature.
3. It may be necessary to assign more than one type to some reason mentions.
4. When broad types overlap or cover various narrow types, reviewers should present data on narrow types whenever practicable. The discussion section of the resulting paper should recommend examination of the relation between reason types, and the limitations section should reflect on the meaning of the data.

Assigning reason types to complex reasons

Our proposal to assign to each mention a broad and a narrow type is unsuitable for extracting complex reasons. Such reasons include more than one premise to which a broad reason type can be applied, and the premises are related. To illustrate:

Requiring sponsors to ensure PTA to the trial drug reduces their incentive to conduct research, which will result in the loss of potential benefits to countries that would otherwise host the research, for example.^{17–19}

Even if one applies two broad types to this passage—incentive and stake-holders' interests—the pair will fail to capture the causal relationship that the authors claim obtain between the reduction of incentive and the loss of potential benefits. Our compromise solution, which sought to keep the analysis manageable and results accessible to decision-makers, was to assign two broad reason types, and to treat the pair as a distinct broad type.

Derive and present results

The results section's key exhibit will be a list of all the types of reasons mentioned in the literature reviewed. This should show, for each type of reason listed, whether it was used by the different publications that mentioned it to argue just for the view in question, or just against, or whether some publications used it to argue for and others against. This list is the

answer to the review question: in our case, the question of which reasons have been given for the views that PTA to the trial drug should or need not be ensured to trial participants. It helps decision-makers and philosophers identify the relevant (and so, ultimately, the strong) reasons, and their most plausible interpretations and uses. If narrow types are finely individuated, this list also helps philosophers to individuate reason concepts. This is because the problem of concept individuation becomes that of how to 'join the dots' between the reasons on the list and of whether different list entries are, in fact, the same.

The abstract of a systematic review of reasons cannot present the complete answer to the review question. The results section of our systematic review of reasons was: 'Of 2060 publications identified, 75 were included. These mentioned reasons based on morality, legality, interests/incentives, or practicality, comprising 36 broad (235 narrow) types of reason. None of the included publications, which included informal reviews and reports by official bodies, mentioned more than 22 broad (59 narrow) types. For many reasons, publications differed about the reason's interpretation, implications and/or persuasiveness. Publications differed also regarding costs, feasibility and legality of PTA' (p. 160).

Key decisions regarding the results are:

1. whether, for each type, to include a count of the number of reason mentions;
2. whether to include a count of the total number of broad types and/or narrow types;
3. how to order the reasons in the list, if not by frequency of occurrence; and
4. whether to list broad reason types and/or narrow reason types.

Regarding (1): before counting the number of mentions of a specific (broad or narrow) reason type, one should first remove reasons repeated within each publication: reasons in which the codes, alleged implication, person expressing attitude and attitude expressed are identical. Care must be taken particularly when counting the number of mentions when these have been used to support different conclusions. Regarding (2): because types could be made narrower or broader, it is advisable to conduct a sensitivity analysis as follows. One uses finely individuated narrow reason types in the initial analysis and counts the number of types, and then merges similar narrow types and recounts the number of types. One then calculates the difference in the counts of types when they are narrowly versus broadly individuated.

Another essential exhibit is a table of the characteristics of included publications. This enables readers to assess the state of the field and identify gaps. It is a good idea if the review also contains a list of all the publications included when a literature is difficult to track down, as in our case. Furthermore, particularly when users of the review are likely to be interested in the positions taken by individual publications, and it takes specialised skill to extract this information, decision-makers may find useful a table that shows, for each publication, the reasons endorsed by the publication, whether the reason was used for and against, the conclusion drawn from the reason and attitude taken by the author, and the publication's overall conclusion. For examples of all these exhibits, see reference.⁸

Table 1 suggests results to derive and present in the exhibits just mentioned. We stress, however, that the choice of results will depend on the review question and literature reviewed. For further results that could be derived and presented see reference.⁸

Limitations to the derivation of results

Depending on the methodology used, the following objections may or may not be relevant. In general, the limitations section of a systematic review of reasons should acknowledge and address the relevant ones.

One likely objection, when total counts of types or of mentions of each type are presented, is that the number of reason types in the literature has little meaning. Types could be narrowed or broadened, and some broad reason types may cover diverse narrow types. Furthermore, there may be similarly good reasons to class a narrow reason type under two different broad types. When we conducted our review, we might have placed the narrow reason type “undue inducement” under the broad reason type “incentive”, because it is incentives that are unduly inductive (or not). Or we might have placed it under “informed consent”, because it is frequently argued that participants unduly induced to participate cannot give valid consent. Furthermore, counts may mislead decision-makers, if they think that more commonly presented reasons are thereby stronger reasons and therefore deserve greater weight in decision-making.

In reply (1): further research should address whether decision-makers tend to think that more commonly presented reasons deserve more weight. If so, it may be favourable to exclude counts from systematic reviews of reasons intended for decision-makers, and to include a discussion of the strength of published reasons. (2) Counts of narrow types are meaningful when (a) narrow types are individuated as finely as possible, (b) the assignment of reason mentions to narrow types is less arbitrary than the assignment of narrow to broad types, and (c) the sensitivity analysis shows that the number of types varies little when narrowly individuated types are broadened. Counts of narrow reasons may still mislead when the literature is unclear; however, the systematic review inherits this limitation from the literature.

Furthermore, unless conclusion types are extremely narrow, counts of reasons will cover reasons that were used to support slightly different conclusions. However, we note that classic systematic reviews face the analogous problem that the studies reviewed addressed slightly different research questions, and that such systematic reviews have developed appropriate strategies. We hypothesise that systematic reviews of reasons will frequently face this problem; research is needed on adapting existing strategies. Although a comparison between traditional systematic reviews and systematic reviews of reasons is beyond the scope of this paper, we would like to point out that the former are also subject to many types of bias.²⁰ In any case, we recommend caution when computing counts and the presentation of qualitative results instead of counts whenever appropriate.

Applications of the model are needed to clarify further obstacles, to identify tactics for minimising external biases (eg, publication bias) and internal or review author-related biases (eg, selection bias and coding bias) and to develop means to evaluate the model.

CONCLUSIONS

Systematic reviews of reasons are needed to guide decisions in medicine and policy. This paper presented a model for writing such reviews. We explained how to apply the model to specific review questions and literatures, identified and addressed various technical and conceptual issues, and considered the

extent to which they can be addressed with a revised qualitative research methodology, appropriate disclaimers and further research. We hope that this detailed and critical adaptation of the systematic review technique to reason-based bioethics will lead to further applications that aim to evaluate the opportunities and limitations of this model and suggest additional modifications.

Acknowledgements The authors would like to thank Reuben Thomas and Leif Wenar.

Funding DS was partly supported by a grant from the German Research Foundation (DFG) (STR 1070/2-1). NS was supported by a research fellowship in biomedical ethics from the Wellcome Trust, grant number 088360.

Competing interests NS is collaborating with the UK's National Research Ethics Service (NRES) to write NRES's first guidance on post-trial access.

Contributors DS and NS contributed substantially to drafting and revising the article. DS and NS approved the final version.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Eden J**, Levit L, Berg A, et al. *Finding What Works in Health Care: Standards for Systematic Reviews*. Institute of Medicine (IOM) Committee on Standards for Systematic Reviews of Comparative Effectiveness Research. Washington, DC: National Academies Press, 2011.
2. **Moher D**, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;**151**:264–9; W64.
3. **Dixon-Woods M**, Agarwal S, Jones D, et al. Synthesising qualitative and quantitative evidence: a review of possible methods. *J Health Serv Res Policy* 2005;**10**:45–53.
4. **Pope C**, Mays N, Popay J. Informing policy making and management in healthcare: the place for synthesis. *Healthcare Policy* 2006;**1**:43–8.
5. **Strech D**, Persad G, Marckmann G, et al. Are physicians willing to ration health care? Conflicting findings in a systematic review of survey research. *Health Policy* 2009;**90**:113–24.
6. **Strech D**, Synofzik M, Marckmann G. Systematic reviews of empirical bioethics. Conceptual challenges and practical recommendations. *J Med Ethics* 2008;**34**:472–7.
7. **Strech D**, Synofzik M, Marckmann G. How physicians allocate scarce resources at the bedside: a systematic review of qualitative studies. *J Med Philos* 2008;**33**:80–99.
8. **Sofaer N**, Strech D. Reasons why post-trial access to trial drugs should, or need not be ensured to research participants: a systematic review. *Public Health Ethics* 2011;**4**:160–84.
9. **McCullough LB**, Coverdale JH, Chervenak FA. Constructing a systematic review for argument-based clinical ethics literature: the example of concealed medications. *J Med Philos* 2007;**32**:65–76.
10. **Sofaer N**, Strech D. The need for systematic reviews of reasons. *Bioethics* 2011. doi:10.1111/j.1467–8519.2011.01858.x. [Epub ahead to print].
11. **CEBM staff**. *Centre for Evidence-Based Medicine: Asking Focused Questions*. 2009. <http://www.cebm.net/index.aspx?o=1036> (accessed 1 Mar 2011).
12. **Mays N**, Pope C, Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *J Health Serv Res Policy* 2005;(10 Suppl 1):6–20.
13. **Droste S**, Dintsios CM, Gerber A. Information on ethical issues in health technology assessment: how and where to find them. *Int J Technol Assess Health Care* 2010;**26**:441–9.
14. **Mulrow CD**. Rationale for systematic reviews. *BMJ* 1994;**309**:597–9.
15. **Popay J**, Rogers A, Williams G. Rationale and standards for the systematic review of qualitative literature in health services research. *Qual Health Res* 1998;**8**:341–51.
16. **Walsh D**, Downe S. Meta-synthesis method for qualitative research: a literature review. *J Adv Nurs* 2005;**50**:204–11.
17. **Cooley DR**. Distributive justice and clinical trials in the Third World. *Theor Med Bioeth* 2001;**22**:151–67.
18. **Nuffield Council on Bioethics Group**. *The Ethics of Research Related to Healthcare in Developing Countries*. London: Nuffield Council on Bioethics, 2002.
19. **Lavery JV**. The obligation to ensure access to beneficial treatments for research participants at the conclusion of clinical trials. In: Emanuel EJ, Grady C, Crouch RA, et al, eds. *The Oxford Textbook of Clinical Research Ethics*. New York: Oxford University Press, 2008:697–710.
20. **Higgins JPT**, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 [updated September 2009]: The Cochrane Collaboration*. 2009. <http://www.cochrane-handbook.org/> (accessed 1 Mar 10).

Table (i): Suggested results to derive (as online supplemental)

Publication characteristics

- Number of publications included in the systematic review
- Range of years in which published and for each year, the number of included publications that were published during the year
- Number and percentage of publications for which PTA was not the main topic
- Number and percentage of English-language publications
- Measure of authors' conflict of interest

Spectrum and incidence of reasons (and conclusions if relevant for the review question)

- The broad types and narrow types of reasons (and conclusions)
- The total number of broad reason types and narrow reason types (excluding repeated reasons). For the purpose of sensitivity analysis count twice: 1) when types are individuated narrowly, and 2) when types are individuated broadly
- For each broad (narrow) reason type, the total number of reason mentions respectively

Reasons endorsed and conclusions drawn by individual publications

- For each included publication, the identity of all the reasons endorsed and any all-things-considered conclusion drawn by the publication. In our first systematic review of reasons, we presented, for example, all reasons endorsed for ensuring PTA and the reasons endorsed for the view that PTA need not be ensured.
- The largest number of broad and narrow types of reasons mentioned by any publication reviewed and by the systematic review itself. Purpose: to illustrate differences between the scope of reasons captured in informal vs. systematic reviews.

ⁱA *MeSH* term is a search term assigned to a Medline article depending on its content rather than vocabulary.

Appendix: How we developed, and how to customize, our model

We first read various methodology publications and systematic reviews, and discussed which data to extract. Then, we created an Excel spreadsheet with row headings for mentions in publications reviewed, and column headings for data types. We decided that some data types should concern the publication, e.g. the citation, whereas other data types should concern individual mentions, e.g. the passage in the publication that expresses the reason. Next, we identified eight publications which we already knew met the eligibility criteria, and which illustrated various reasons and challenges to analysis.

Each author then independently completed the spreadsheet for publication 1. Discrepancies between spreadsheets were identified, discussed and resolved; consequently, certain column headings were changed. We next re-coded publication 1 independently and repeated the process, and then independently completed the resulting spreadsheet for publications 2-3. The entire process was repeated until coding additional publications resulted in no further changes to

headings and until discrepancies between corresponding cells were easily resolved. Frequent discussions identified and addressed problems.

Throughout the process, we kept a list of questions answerable only after coding further publications. One such question was whether the same code should be applied *both* to reasons based on incentives for people to participate in research *and* to reasons based on the need to reward participants. We added questions and answers as data extraction progressed.

Authors should use the above process with an initial spreadsheet that uses the types of data we chose (listed below) as column headings. They should then modify the headings based on the content of the literature included in their systematic review of reasons and according to their specific review question and literature to review.